

# UC Davis

## UC Davis Previously Published Works

### Title

Complete genome sequence of Eggerthella lenta type strain (IPP VPI 0255).

### Permalink

<https://escholarship.org/uc/item/0qh29089>

### Journal

Standards in genomic sciences, 1(2)

### ISSN

1944-3277

### Authors

Saunders, Elizabeth  
Pukall, Rüdiger  
Abt, Birte  
et al.

### Publication Date

2009-09-01

### DOI

10.4056/sigs.33592

Peer reviewed

## Complete genome sequence of *Eggerthella lenta* type strain (VPI 0255<sup>T</sup>)

Elizabeth Saunders<sup>1</sup>, Rüdiger Pukall<sup>2</sup>, Birte Abt<sup>2</sup>, Alla Lapidus<sup>1</sup>, Tijana Glavina Del Rio<sup>1</sup>, Alex Copeland<sup>1</sup>, Hope Tice<sup>1</sup>, Jan-Fang Cheng<sup>1</sup>, Susan Lucas<sup>1</sup>, Feng Chen<sup>1</sup>, Matt Nolan<sup>1</sup>, David Bruce<sup>1,3</sup>, Lynne Goodwin<sup>1,3</sup>, Sam Pitluck<sup>1</sup>, Natalia Ivanova<sup>1</sup>, Konstantinos Mavromatis<sup>1</sup>, Galina Ovchinnikova<sup>1</sup>, Amrita Pati<sup>1</sup>, Amy Chen<sup>4</sup>, Krishna Palaniappan<sup>4</sup>, Miriam Land<sup>1,5</sup>, Loren Hauser<sup>1,5</sup>, Yun-Juan Chang<sup>1,5</sup>, Cynthia D. Jeffries<sup>1,5</sup>, Patrick Chain<sup>1,6</sup>, Linda Meincke<sup>1,3</sup>, David Sims<sup>1,3</sup>, Thomas Brettin<sup>1,3</sup>, John C. Detter<sup>1,3</sup>, Markus Göker<sup>2</sup>, Jim Bristow<sup>1</sup>, Jonathan A. Eisen<sup>1,7</sup>, Victor Markowitz<sup>4</sup>, Philip Hugenholtz<sup>1</sup>, Nikos C. Kyrpides<sup>1</sup>, Hans-Peter Klenk<sup>2</sup>, and Cliff Han<sup>1,3\*</sup>

<sup>1</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>2</sup> DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

<sup>3</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

<sup>4</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>5</sup> Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>6</sup> Lawrence Livermore National Laboratory, Livermore, California, USA

<sup>7</sup> University of California Davis Genome Center, Davis, California, USA

\*Corresponding author: Cliff Han

**Keywords:** mesophile, anaerobic, human intestinal microflora, pathogenic, bacteremia, Gram-positive, *Coriobacteriaceae*

---

*Eggerthella lenta* (Eggerth 1935) Wade *et al.* 1999, emended Würdemann *et al.* 2009 is the type species of the genus *Eggerthella*, which belongs to the actinobacterial family *Coriobacteriaceae*. *E. lenta* is a Gram-positive, non-motile, non-sporulating pathogenic bacterium that can cause severe bacteremia. The strain described in this study has been isolated from a rectal tumor in 1935. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of the genus *Eggerthella*, and the 3,632,260 bp long single replicon genome with its 3123 protein-coding and 58 RNA genes is part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

---

### Introduction

Strain VPI 0255<sup>T</sup> (= DSM 2243 = ATCC 25559 = JCM 9979) is the type strain of the species *Eggerthella lenta*, which was first described in 1935 by Eggerth as '*Bacteroides lentus*' [1], later in 1938 renamed by Prévot in '*Eubacterium lentum*' [2], and was also known under the synonym '*Pseudobacterium lentum*' Krasil'nikov 1949 [3]. The strain has been described in detail by Moore *et al.* in 1971 [4]. Based on 16S rRNA sequence divergence and the presence of unique phenotypic characters the strain was then transferred to the new genus *Eggerthella* as *E. lenta* (Kageyama *et al.* 1999, Wade *et al.* 1999 [5,6] In 2004 two novel *Eggerthella* species, *E. hongkongensis* and *E. sinensis* were characterized and described in addition [7].

Recently, *E. hongkongensis* was reclassified as *Paraeggerthella hongkongensis* [8]. Although the two *Eggerthella* species and *P. hongkongensis* are part of the human gut flora, they can be the agent of severe bacteremia. So far the pathogenic potential of the genera are poorly analyzed [7]. Here we present a summary classification and a set of features for *E. lenta* VPI 0255<sup>T</sup>, together with the description of the complete genomic sequencing and annotation.

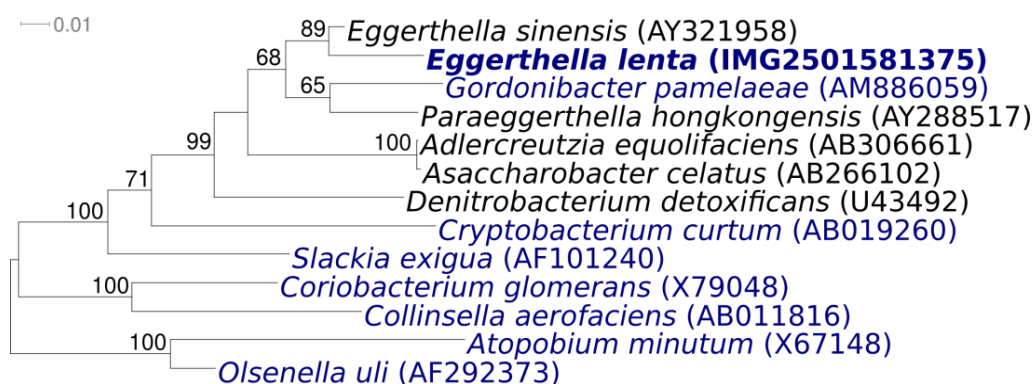
### Classification and features

Members of the species *E. lenta* have been isolated from several abscesses, from appendix tissues, peritoneal fluid and intestinal tumors. The organ-

ism is often involved in mixed infections with less fastidious bacteria. Difficulties in cultivation and identification are probably the reason why bacteremia caused by *Eggerthella* is rarely reported. Half of the cases of *Eggerthella* bacteremia are induced by the two novel species: *E. sinensis* and *P. hongkongensis* [7]. Stinear *et al.* described an isolate (AF304434) from human feces resembling *E. lenta* (98% identity) that carries an enterococcal *vanB* resistance locus probably received via lateral gene transfer or as a result of genetic mutations [9]. Clavel *et al.* investigated the occurrence and activity of dietary lignans activating bacterial communities in human feces and identified an *E. lenta* strain (AY937380) with 98.2% sequence similarity to strain VPI 0255<sup>T</sup> [10]. Lignans are a class of phytoestrogen which can be metabolized to the biologically active enterolignans, enterodiols and enterolactone. The human intestinal microbiota is essential for the conversion of the dietary lignans *e.g.* secoisolariciresinol diglucoside via

secoisolariciresinol (SECO) to the enterolignans. Clavel and co-workers also reported that the dehydroxylation of SECO is catalyzed by *Eggerthella lenta* [11]. Based on 16S rRNA gene sequence analyses another five uncultured clones with 99% identity to *E. lenta* were reported at the NCBI BLAST server (status June 2009). These clones were derived from the analyses of feces samples from humans. *e.g.* associated with obesity [12,13], but also from marine metagenomes [14]

Figure 1 shows the phylogenetic neighborhood of *E. lenta* strain VPI 0255<sup>T</sup> in a 16S rRNA based tree. The sequences of the three identical copies of the 16S rRNA gene in the genome differ by three nucleotides from the previously published 16S rRNA sequence generated from ATCC 25559 (AF292375). The slight difference between the genome data and the reported 16S rRNA gene sequence is most likely due to sequencing errors in the previously reported sequence data.



**Figure 1.** Phylogenetic tree of *E. lenta* strain VPI 0255<sup>T</sup> and all type strains of the genus *Eggerthella* as well as the type strains from all other genera of the family *Coriobacteriaceae* inferred from 1,373 aligned characters [15,16] of the 16S rRNA gene under the maximum likelihood criterion [17]. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1,000 bootstrap replicates if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [18] are shown in blue, published genomes in bold, including two of which are reported in this issue of *SIGS* [19,20]

*E. lenta* strain VPI 0255<sup>T</sup> was originally isolated from a rectal tumor and described as Gram-positive, non-motile and non-sporulating (Table 1) [1]. Cells are rod shaped and occur singly or in long chains up to 20 elements (Figure 2). The cell size and morphology varies depending on the substrate and the age of the culture. Surface colonies were described as circular to slightly scalloped, convex, shiny, gray and translucent. *E. lenta* is ob-

ligately anaerobic and its optimal growth temperature is 37° C [4]. Growth is stimulated by arginine. The existence of the arginine dihydrolase pathway as an important energy source was described by Sperry and Wilkens in 1976 [26]. *E. lenta* is asaccharolytic [4,26,29], Gelatin is not liquefied, aesculin is not hydrolyzed and nitrate is reduced [29]. *E. lenta* is bile-resistant and primarily found in human feces [6].

## Chemotaxonomy

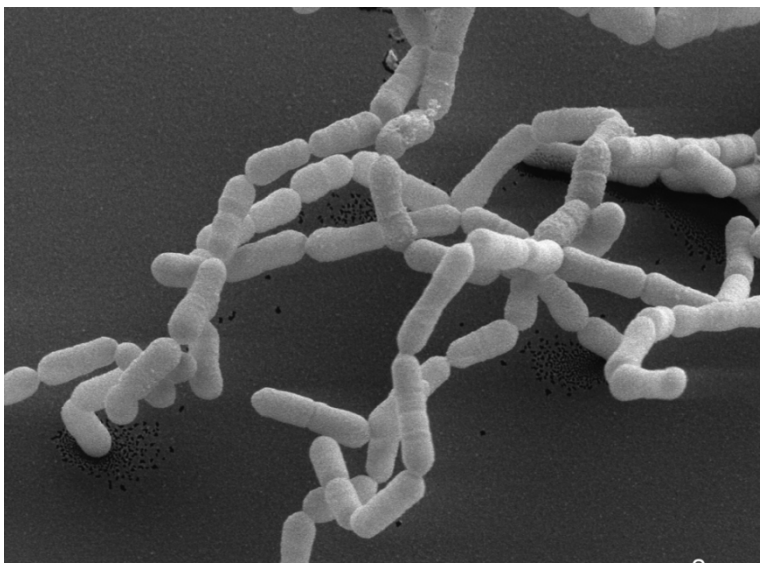
The cell wall of *E. lenta* strain VPI 0255<sup>T</sup> contains A1γ-type peptidoglycan glutamic acid occurred in D-form and diaminopimelic acid in *meso* configuration. Mycolic acids and teichonic acids were not reported. Strain VPI 0255<sup>T</sup> contains menaquinone MK-6 as the major respiratory lipoquinone (63.7%) and a lower amount of the methylmenaquinone

MMK-6 (36.3%) [8,29,31]. As the predominant fatty acids the unbranched saturated 16:0 DMA (29.4%) and the monounsaturated fatty acid 18:1w9c (22.0%) were identified [5,6]. Polar lipids consist of two phospholipids, phosphatidylglycerol and diphosphatidylglycerol, and four glycolipids GL1-GL4 [8].

**Table 1.** Classification and general features of *B. cavernae* HKI 0122<sup>T</sup> in accordance with the MIGS recommendations [21]

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [22]
		Phylum <i>Actinobacteria</i>	TAS [23]
		Class <i>Actinobacteria</i>	TAS [24]
		Order <i>Coriobacteriales</i>	TAS [24]
	Classification	Suborder " <i>Coriobacterineae</i> "	TAS [25]
		Family <i>Coriobacteriaceae</i>	TAS [24]
		Genus <i>Eggerthella</i>	TAS [6]
		Species <i>Eggerthella lenta</i>	TAS [6]
		Type strain VPI 0255	
	Gram stain	positive	TAS [1,4]
	Cell shape	rods, single or arranged in pairs and chains	TAS [1,4]
	Motility	non-motile	TAS [1,4]
	Sporulation	non-sporulating	TAS [1,4]
	Temperature range	mesophile	TAS [4]
	Optimum temperature	37°C	TAS [4]
	Salinity	6.5% NaCl, poor to moderate growth	TAS [4]
MIGS-22	Oxygen requirement	anaerobic	TAS [1,4]
	Carbon source	arginine	TAS [24,26]
	Energy source	arginine	TAS [26]
MIGS-6	Habitat	blood, human intestinal microflora	TAS [1,7]
MIGS-15	Biotic relationship	free living	NAS
MIGS-14	Pathogenicity	bacteremia	TAS [27]
	Biosafety level	2	TAS [28]
	Isolation	rectal tumor	TAS [1,29]
MIGS-4	Geographic location	not reported	
MIGS-5	Sample collection time	1938	TAS [1]
MIGS-4.1	Latitude – Longitude	not reported	
MIGS-4.2			
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	not reported	

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [30]. If the evidence code is IDA, then the property was directly observed for a living isolate by one of the authors, or an expert or reputable institution mentioned in the acknowledgements.



**Figure 2.** Scanning electron micrograph of *E. lenta* VPI 0255<sup>T</sup> (Manfred Rohde, Helmholtz Centre for Infection Biology, Braunschweig)

## Genome sequencing and annotation

### Genome project history

This organism was selected for sequencing on the basis of each phylogenetic position, and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project. The genome project is deposited in the Genome OnLine Database [18] and the complete genome sequence in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

### Growth conditions and DNA isolation

*E. lenta* strain VPI 0255<sup>T</sup>, DSM 2243, was grown anaerobically in DSMZ medium 209 (*Eubacterium lentum* Medium [32]) at 37°C. DNA was isolated from 1-1.5 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol without modifications.

**Table 2.** Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Three genomic libraries: two Sanger libraries - 8 kb pMCL200 and fosmid pcc1Fos – and one 454 pyrosequence standard library
MIGS-29	Sequencing platforms	ABI3730, 454 GS FLX
MIGS-31.2	Sequencing coverage	10.2× Sanger; 25.3× pyrosequence
MIGS-30	Assemblers	Newbler version 1.1.02.15, phrap
MIGS-32	Gene calling method	Prodigal, GenePRIMP
	Genbank ID	CP001726
	Genbank Date of Release	September 9, 2009
	GOLD ID	Gc01054
	NCBI project ID	21093
	Database: IMG-GEBA	2501533210
MIGS-13	Source material identifier	DSM 2243
	Project relevance	Tree of Life, GEBA

## Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website. 454 Pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 4,901 overlapping fragments of 1,000 bp and entered into the assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and to adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the parallel phrap assembler (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher or transposon bombing of bridging clones [33]. Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification. A total of 358 Sanger finishing reads were produced to close gaps, to resolve repetitive regions, and to raise the quality of the finished sequence. The final assembly consists of 39,464 Sanger and 471,609 pyrosequence (454) reads. Together all sequence types provided 35.5x coverage of the genome. The error rate of the completed genome sequence is less than 1 in 100,000.

## Genome annotation

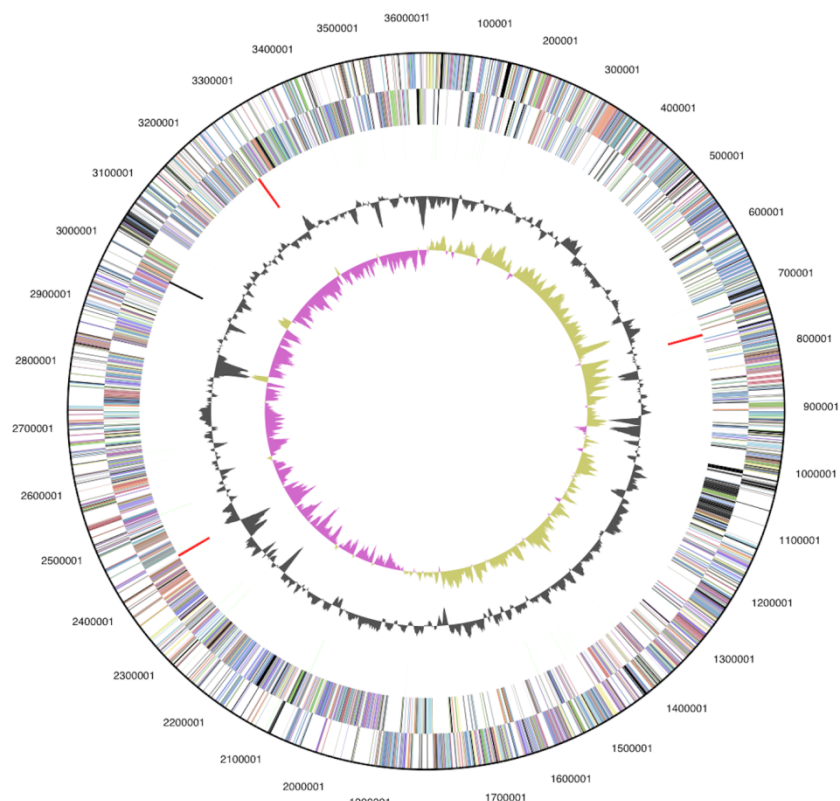
Genes were identified using Prodigal [34] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [35]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes Expert Review (IMG-ER) platform [36].

## Genome properties

The genome is 3,632,260 bp long and comprises one main circular chromosome with a 64.2% GC content (Table 3 and Figure 3). Of the 3,181 genes predicted, 3,123 were protein coding genes, and 58 RNAs. 53 pseudogenes were also identified. A majority of the genes (70.9%) were assigned with a putative function while the remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 3. The distribution of genes into COGs functional categories is presented in Table 4.

**Table 3.** Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	3,632,260	100.00%
DNA Coding region (bp)	3,211,405	88.41%
DNA G+C content (bp)	2,322,078	64.20%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	3,181	100.00%
RNA genes	58	1.67%
rRNA operons	3	
Protein-coding genes	3,123	98.18%
Pseudo genes	53	1.67%
Genes with function prediction	2,255	70.89%
Genes in paralog clusters	629	19.77%
Genes assigned to COGs	2285	71.83%
Genes assigned Pfam domains	2316	72.81%
Genes with signal peptides	781	24.55%
Genes with transmembrane helices	990	31.12%
CRISPR repeats	1	



**Figure 3.** Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

**Table 4.** Number of genes associated with the general COG functional categories

Code	Value	%age	Description
J	142	4.5	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	310	9.9	Transcription
L	130	4.2	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	25	0.8	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	80	2.6	Defense mechanisms
T	201	6.4	Signal transduction mechanisms
M	129	4.1	Cell wall/membrane biogenesis
N	13	0.4	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	51	1.6	Intracellular trafficking and secretion
O	81	2.6	Posttranslational modification, protein turnover, chaperones
C	293	9.4	Energy production and conversion
G	79	2.5	Carbohydrate transport and metabolism
E	180	5.8	Amino acid transport and metabolism



**Table 4.** Number of genes associated with the general COG functional categories (cont.)

Code	Value	%age	Description
F	60	1.9	Nucleotide transport and metabolism
H	89	2.8	Coenzyme transport and metabolism
I	69	2.2	Lipid transport and metabolism
P	132	4.2	Inorganic ion transport and metabolism
Q	32	1.0	Secondary metabolites biosynthesis, transport and catabolism
R	262	8.4	General function prediction only
S	195	6.2	Function unknown
-	838	26.8	Not in COGs

## Acknowledgements

We would like to gratefully acknowledge the help of Gabriele Gehrich-Schröter for growing *E. lenta* cultures and Susanne Schneider for DNA extraction and quality analysis (both at DSMZ). This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California,

Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, as well as German Research Foundation (DFG) INST 599/1-1.

## References

1. Eggerth AH. The Gram-positive non-spore-bearing anaerobic Bacilli of human feces. *J Bacteriol* 1935; **30**:277-299. [PubMed](#)
2. Prévot AR. Études de systématique bactérienne. III. Invalidité du genre *Bacteroides* Castellani et Chalmers. Démembrement et reclassification. *Ann Inst Pasteur (Paris)* 1938; **60**:295.
3. Krasil'nikov NA. Guide to the Bacteria and Actinomycetes [Opredelitel' Bakterii i Actinomycetov]. *Akad. Nauk SSSR, Moscow* 1949; 1-830.
4. Moore WEC, Cato EP, Holdeman LV. *Eubacterium lentum* (Eggerth) Prévot 1938: Emendation of description and designation of the neotype strain. *Int J Syst Bacteriol* 1971; **21**:299-303.
5. Kageyama A, Benno Y, Nakase T. Phylogenetic evidence for transfer of *Eubacterium lentum* to the genus *Eggerthella* as *Eggerthella lenta* gen. nov., comb. nov. *Int J Syst Bacteriol* 1999; **49**:1725-1732. [PubMed](#)
6. Wade WG, Downes J, Dymock D, Hiom SJ, Weightman AJ, Dewhirst FE, Paster BJ, Tzellas N, Coleman B. The family *Coriobacteriaceae*: reclassification of *Eubacterium exiguum* (Poco et al. 1996) and *Peptostreptococcus heliotrinireducens* (Lanigan 1976) as *Slackia exigua* gen. nov., comb. nov. and *Slackia heliotrinireducens* gen. nov., comb. nov., and *Eubacterium lentum* (Prevot 1938) as *Eggerthella lenta* gen. nov., comb. nov. *Int J Syst Bacteriol* 1999; **49**:595-600. [PubMed](#)
7. Lau SK, Woo PC, Woo GK, Fung AM, Wong MK, Chan KM, Tam DM, Yuen KY. *Eggerthella hongkongensis* sp. nov. and *Eggerthella sinensis* sp. nov., two novel *Eggerthella* species, account for half of the cases of *Eggerthella* bacteremia. *Diagn Microbiol Infect Dis* 2004; **49**:255-263 [PubMed](#) doi:10.1016/j.diagmicrobio.2004.04.012
8. Würdemann D, Tindall BJ, Pukall R, Lünsdorf H, Strömpl C, Namuth T, Nahrstedt H, Wos-Oxley M, Ott S, Schreiber S, et al. *Gordonibacter pame-laeae* gen. nov., sp. nov., a new member of the *Coriobacteriaceae* isolated from a patient with Crohn's disease, and reclassification of *Eggerthella hongkongensis* Lau et al. 2006 as *Paraeggerthella hongkongensis* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 2009; **59**:1405-1415. [PubMed](#) doi:10.1099/ijs.0.005900-0
9. Stinear TP, Olden DC, Johnson PDR, Davies JK, Grayson L. Enterococcal *vanB* resistance locus in anaerobic bacteria in human faeces. *Lancet* 2001; **357**:855-856. [PubMed](#) doi:10.1016/S0140-6736(00)04206-9
10. Clavel T, Henderson G, Alpert CA, Philippe C, Rigottier-Gois L, Doré J, Blaut M. Intestinal bacterial communities that produce active estrogen-like compounds enterodiol and enterolactone in humans. *Appl Environ Microbiol* 2005; **71**:6077-6085. [PubMed](#) doi:10.1128/AEM.71.10.6077-6085.2005



11. Clavel T, Borrmann D, Braune A, Dore J, Blaut M. Occurrence and activity of human intestinal bacteria involved in the conversion of dietary lignans. *Anaerobe* 2006; **12**:140-147.  
[PubMed doi:10.1016/j.anaerobe.2005.11.002](#)
12. Ley RE, Turnbaugh PJ, Klein S, Gordon JL. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006; **444**:1022-1023.  
[PubMed doi:10.1038/4441022a](#)
13. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbioms. *DNA Res* 2007; **14**:169-181.  
[PubMed doi:10.1093/dnares/dsm018](#)
14. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen J, Nelson KE, Nelson W, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; **304**:66-74.  
[PubMed doi:10.1126/science.1093857](#)
15. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464.  
[PubMed doi:10.1093/bioinformatics/18.3.452](#)
16. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552.  
[PubMed](#)
17. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771.  
[PubMed doi:10.1080/10635150802429642](#)
18. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes OnLine Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479.  
[PubMed doi:10.1093/nar/gkm884](#)
19. Mavrommatis K, Pukall R, Rohde C, Chen F, Sims D, Brettin T, Kuske C, Detter JC, Han C, Lapidus A, et al. Complete genome sequence of *Cryptobacterium curtum* type strain (12-3T). *Stand Genomic Sci* 2009; **1**: 93-100 .  
[doi:10.4056/sigs.15195](#)
20. Copeland A, Sikorski J, Lapidus A, Nolan M, Galvina Del Rio T, Lucas S, Chen F, Tice H, Pitluck S, Cheng JF, et al. Complete genome sequence of *Atopobium parvulum* type strain (IPP 1246T). *Stand Genomic Sci* 2009; **1**: 166-173-8.  
[doi:10.4056/sigs.15195](#)
21. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547.  
[PubMed doi:10.1038/nbt1360](#)
22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**: 4576-4579.  
[PubMed doi:10.1073/pnas.87.12.4576](#)
23. Garrity GM, Holt J. Taxonomic Outline of the Archaea and Bacteria. Bergey's Manual of Systematic Bacteriology, 2<sup>nd</sup> Ed. In: G.Garrity GM, Boone DR, Castenholz RW Eds. Vol 1 The *Archaea*, Deeply Branching and Phototrophic *Bacteria*. 2001 pp. 155-166
24. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a New Hierarchic Classification System, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997; **47**:479-491.
25. Garrity GM, Bell JA, Lilburn T. In: Garrity GM, Boone DR, Castenholz RW (2001) Taxonomic outline of the Procaryotes. Bergey's Manual of Systematic Bacteriology 1-39.
26. Sperry JF, Wilkins TD. Arginine, a growth-limiting factor for *Eubacterium lentum*. *J Bacteriol* 1976; **127**:780-784. [PubMed](#)
27. Lau SKP, Woo PCY, Fung AMY, Chan K, Woo GKS, Yuen K. Anaerobic, non-sporulating, Gram-positive bacilli bacteremia characterized by 16S rRNA gene sequencing. *J Med Microbiol* 2004; **53**:1247-1253.  
[PubMed doi:10.1099/jmm.0.45803-0](#)
28. Anonymous. Biological Agents: Technical rules for biological agents www.baua.de TRBA 466.
29. Maruo T, Sakamoto M, Ito C, Toda T, Benno Y. *Adlercreutzia equolifaciens* gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus *Eggerthella*. *Int J Syst Evol Microbiol* 2008; **58**:1221-1227. [PubMed doi:10.1099/ijs.0.65404-0](#)
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29.  
[PubMed doi:10.1038/75556](#)

31. Fernandez F, Collins MD. Vitamin K composition of anaerobic gut bacteria. *FEMS Microbiol Lett* 1987; **41**:175-180. doi:[10.1111/j.1574-6968.1987.tb02191.x](https://doi.org/10.1111/j.1574-6968.1987.tb02191.x)
32. List of media used at DSMZ for cell growth: [http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php)
33. Sims D, Brettin T, Detter JC, Han C, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Chen F, Lucas S, *et al.* Complete genome of *Kytococcus sedentarius* type strain (541<sup>T</sup>). *Stand Genomic Sci* 2009;1:12-20 doi:[10.4056/sigs.761](https://doi.org/10.4056/sigs.761)
34. Anonymous. Prodigal Prokaryotic Dynamic Programming Genefinding Algorithm. Oak Ridge National Laboratory and University of Tennessee 2009 <http://compbio.ornl.gov/prodigal>
35. Pati A, Ivanova N, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. (Submitted).
36. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. Expert IMG ER: A system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. PubMed doi:[10.1093/bioinformatics/btp393](https://doi.org/10.1093/bioinformatics/btp393)